limited as the N3C data did not clarify COVID-19 infection history in those undergoing CPETs.

This study is limited by its lack of data on inpatient versus outpatient status for these tests and the lack of data on COVID-19 infection status for the included sites. However, these would not change overall test numbers. This study also did not evaluate clinical outcomes affected by this reduced testing, an important area for future study. As such, this study opens the door for many speculations as we continue analyzing trends of subsequent waves of COVID-19 infections and variants. These data and subsequent studies based on similar trends may inform the necessity of performing these tests as future studies explore correlations to clinical outcomes of pulmonary disease. ■

**Author disclosures** are available with the text of this letter at www.atsjournals.org.

**Acknowledgment:** The analyses described in this publication were conducted with data or tools accessed through the National Center for Advancing Translational Sciences (NCATS) N3C Data Enclave (https://covid.cd2h.org/enclave) and supported by NCATS U24 TR002306. This research was possible because of the patients whose information is included within the data from participating organizations (https://ncats.nih.gov/n3c/resources/data-contribution/data-transfer-agreement-signatories) and scientists who have contributed to the ongoing development of this community resource (https://doi.org/10.1093/jamia/ocaa196). The project described was supported by the National Institute of General Medical Sciences, 5U54GM104942-04. The content is solely the responsibility of the authors and does not necessarily represent the official views of the U.S. National Institutes of Health. Authorship was determined using International Committee of Medical Journal Editors (ICMJE) recommendation.

Justin T. Call, M.D.
Augustine S. Lee, M.D.
*Mayo Clinic*
*Jacksonville, Florida*

Melissa A. Haendel, Ph.D.
*University of Colorado Anschutz Medical Campus*
*Aurora, Colorado*

Christopher G. Chute, M.D.
*Johns Hopkins University*
*Baltimore, Maryland*

Scott A. Helgeson, M.D.*
*Mayo Clinic*
*Jacksonville, Florida*

ORCID ID: 0000-0001-7590-2293 (S.A.H.).

*Corresponding author (e-mail: helgeson.scott@mayo.edu).

## References

1 Jain S, Santhosh L. On treatments and tests deferred: preparing for collateral damage from COVID-19. *Ann Am Thorac Soc* 2020;17:1371–1373.
2 Kouri A, Gupta S, Yadollahi A, Ryan CM, Gershon AS, To T, *et al.* Addressing reduced laboratory-based pulmonary function testing during a pandemic. *Chest* 2020;158:2502–2510.
3 Ferguson N, Laydon D, Nedjati Gilani G, Imai N, Ainslie K, Baguelin M, *et al.* Report 9: impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. 2020 [accessed 2020 Mar 16]. Available from: https://spiral.imperial.ac.uk/handle/10044/1/77482.
4 Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* 2020;368:860–868.
5 Chang SH, Jiang J, Kon ZN, Williams DM, Geraci TC, Smith DE, *et al.* Safety and efficacy of bronchoscopy in critically ill patients with coronavirus disease 2019. *Chest* 2021;159:870–872.
6 Torrego A, Pajares V, Fernández-Arias C, Vera P, Mancebo J. Bronchoscopy in patients with COVID-19 with invasive mechanical ventilation: a single-center experience. *Am J Respir Crit Care Med* 2020;202:284–287.
7 Gao Y, Chen R, Geng Q, Mo X, Zhan C, Jian W, *et al.* Cardiopulmonary exercise testing might be helpful for interpretation of impaired pulmonary function in recovered COVID-19 patients. *Eur Respir J* 2021;57:2004265.
8 Davis HE, Assaf GS, McCorkell L, Wei H, Low RJ, Re'em Y, *et al.* Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *EClinicalMedicine* 2021;38:101019.



Check for updates

## External Validation of a Risk Score for Daily Prediction of Atrial Fibrillation among Critically Ill Patients with Sepsis

*To the Editor*:

Atrial fibrillation (AF) occurs frequently among patients with sepsis (1–3) and is associated with short- and long-term morbidity and mortality (1, 4). Predicting which patients will develop AF during sepsis can enrich trials that seek to study and prevent AF in critical illness and may aid management decisions for clinicians. One prior risk score has been developed to predict new-onset AF among critically ill patients with sepsis (5), but this has not been validated outside of the original publication. We sought to externally validate performance of AF prediction in a cohort of critically ill patients with sepsis.

Supported by U.S. National Heart, Lung, and Blood Institute grant R01HL136660.

*Author Contributions*: J.M.R., N.A.B., and A.J.W. were involved in conceptual design, data acquisition, analysis, interpretation, and manuscript preparation. E.K.Q. was involved in data acquisition. J.M.R., N.A.B., E.K.Q., K.H.C., D.D.M., and A.J.W. were involved in revising the manuscript for important intellectual content, approved the final version submitted for publication, and are in agreement with the accuracy and integrity of the submitted work.

### Methods

The transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) checklist was used to design and conduct this study (6). We used the Medical Information Mart for Intensive Care III data set (7), which consists of data from ~60,000 intensive care unit (ICU) admissions at a single U.S. tertiary-care hospital. We identified adult patients (≥18 yr) admitted to the ICU with sepsis. Sepsis was defined by an *International Classification*

**Table 1.** Characteristics of patients admitted to the ICU with sepsis

| | External Validation Cohort | | Initial Derivation Cohort | |
|---|---|---|---|---|
| | No AF (*n* = 7,290) | Developed AF (*n* = 554) | No AF (*n* = 1,364) | Developed AF (*n* = 418) |
| Patient characteristics | | | | |
| Age, yr | 67 (54–79) | 76 (66–83) | 57 (45–67) | 66 (59–73) |
| Sex, female | 3,406 (47) | 239 (43) | 593 (43) | 155 (37) |
| Race/ethnicity | | | | |
| Black | 677 (9) | 32 (6) | — | — |
| White | 5,202 (71) | 420 (76) | 1,173 (86) | 381 (91) |
| Hispanic | 222 (3) | 8 (1) | — | — |
| Asian | 186 (3) | 9 (2) | — | — |
| Other | 217 (3) | 10 (2) | — | — |
| Unknown | 786 (11) | 75 (14) | — | — |
| Obesity (BMI > 30 kg/m$^2$)* | 1,746 (33) | 167 (37) | 214 (16) | 84 (20) |
| Admission characteristics | | | | |
| Immunosuppressed† | 253 (4) | 15 (3) | 335 (25) | 125 (30) |
| Use of vasopressors or inotropes‡ | 2,501 (34) | 261 (47) | 787 (58) | 330 (79) |
| Renal failure§ | 3,094 (42) | 279 (50) | 523 (38) | 186 (44) |
| Serum $K^+$* | 4.3 (3.5–5) | 4.5 (3.6–5.1) | 4.4 (4.1–4.9) | 4.6 (4.2–5.1) |
| Highest $F_{IO_2}$* | 0.5 (0.3–0.7) | 0.6 (0.4–0.9) | — | — |
| Inflammation*‖ | | | | |
| None | 4,170 (58) | 308 (56) | 327 (24) | 85 (20) |
| Moderate | 2,657 (37) | 210 (38) | 529 (39) | 145 (35) |
| Severe | 426 (6) | 30 (6) | 508 (37) | 188 (45) |
| Outcome | | | | |
| ICU mortality | 1,411 (19) | 185 (33) | 191 (14) | 120 (29) |
| 90-d mortality | — | — | 409 (30) | 195 (47) |
| 1-yr mortality | — | — | 546 (40) | 254 (61) |

*Definition of abbreviations*: AF = atrial fibrillation; BMI = body mass index; $F_{IO_2}$ = fraction of inspired oxygen; ICU = intensive care unit; $K^+$ = potassium.

Data are expressed as *n* (%) or median (interquartile range). Percentages may not sum to 100% because of rounding.

*Covariate missingness in validation cohort (% missing): obesity (24%), serum $K^+$ (6%), inflammation (5%), and $F_{IO_2}$ (3%).

†Derivation cohort definition: prior use of corticosteroids in high doses (equivalent to prednisolone >75 mg/d for at least 1 wk), acquired immunodeficiency syndrome, current use of immunosuppressive drugs, current use of antineoplastic drugs, recent hematologic malignancy, and any documented humoral or cellular deficiency. Validation cohort definition: human immunodeficiency virus, acquired immunodeficiency syndrome, and hematologic malignancy.

‡Medications assessed in the derivation cohort: norepinephrine, epinephrine, dobutamine, and milrinone. Medications assessed in the validation cohort: norepinephrine, epinephrine, dobutamine, milrinone, vasopressin, phenylephrine, and dopamine.

§Creatinine ⩾1.36 mg/dl or use of renal replacement therapy.

‖Moderate inflammation: white blood cell count 15–29.9 × $10^9$/L or C-reactive protein 70–149.9 mg/L. Severe inflammation: white blood cell count ⩾30 × $10^9$/L or C-reactive protein ⩾150 mg/L.

*of Disease, Ninth Revision* code for sepsis or a combination of International Classification of Disease, Ninth Revision codes for infection and organ dysfunction (8). The initial admission was used for patients with multiple ICU stays. Individuals with preexisting AF, or AF documented before ICU admission, were excluded.

The primary outcome of new AF occurrence was assessed by hourly nurse-charted heart rhythms (9) and defined as any occurrence of AF in the ensuing 24 hours. Definitions for model variables were similar to the original prediction model (5) with the exception of immunosuppression (unavailable data fields for administration of steroids and outpatient immunosuppressive medications). Data were collected for the first 7 days of ICU admission to harmonize with the reference study. Time-varying explanatory variables were aggregated over each 24-hour period: potassium farthest from 4 mmol/L, highest fraction of inspired oxygen, and highest degree of inflammation. The last observation carried forward was used to impute missing time-varying variables. Multiple imputation with chained equations was used to impute missing baseline covariates across 10 imputed data sets (10). Sensitivity analyses were performed using complete cases and limiting to patients aged ⩾40 years; subgroup analyses assessed performance across ICU types.

We assessed external validity of the original prediction model (5) in predicting AF occurrence in the ensuing 24 hours in a novel cohort, evaluating the same performance measures as the initial study: discrimination was assessed using C-statistic, goodness of fit with a modified Hosmer–Lemeshow (HL) test accounting for large sample size (11), and calibration by plotting the observed versus predicted risk of AF, and using an integrated calibration index (ICI) (12) that calculates the weighted difference between observed and predicted AF rate. In addition, we determined positive predictive value (PPV) for the model at the optimal cut-point of sensitivity and specificity based on Youden's index (13). We then revised the model in a training cohort (random 75% subset of patients in our data set), and test cohort (remaining 25% of patients), using the same covariates but recalculating intercept and β estimates. All statistical analyses were performed with R version 3.6.1 in R-studio version

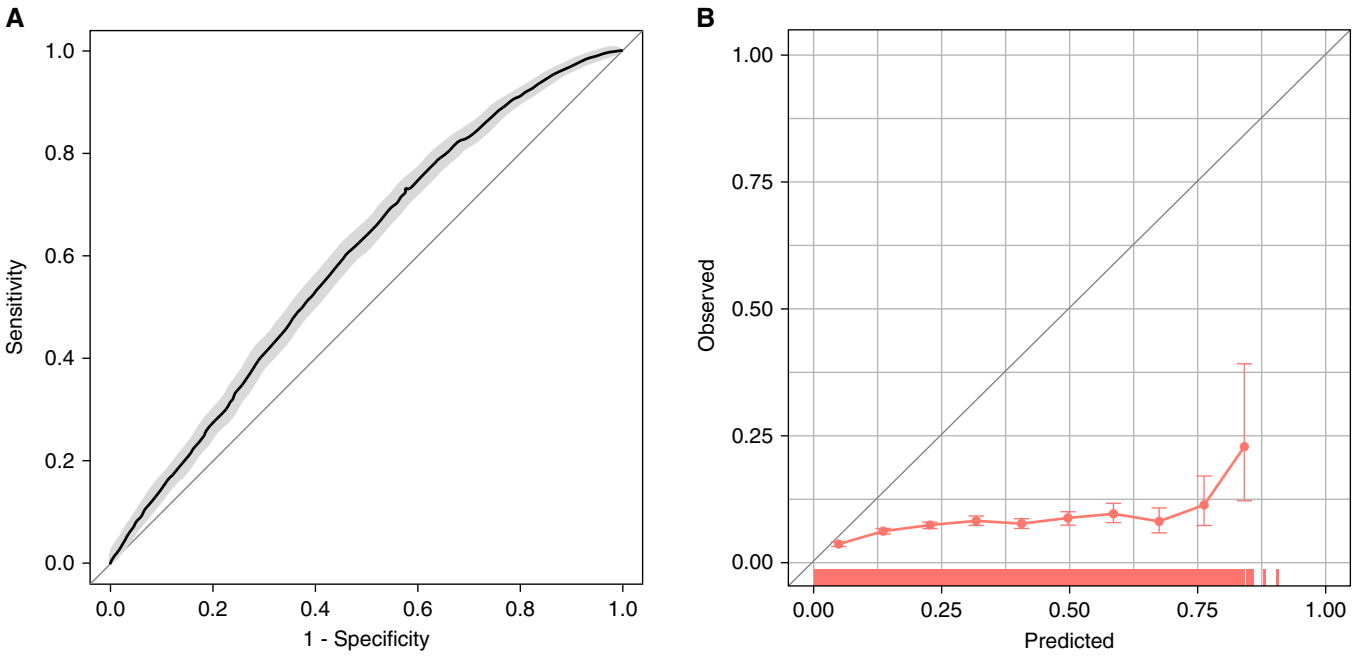


**Figure 1.** Test performance of an existing prediction model for new-onset atrial fibrillation in an external validation cohort. (*A*) Receiver operating characteristic curve for prediction of new-onset atrial fibrillation in the next 24 hours. Area under the receiver operating characteristic curve = 0.598 (95% confidence interval [CI], 0.587–0.609). (*B*) Calibration curve for prediction of new-onset atrial fibrillation in the next 24 hours. Each circle represents a grouping of patients by predicted probability of atrial fibrillation and the corresponding observed occurrence of atrial fibrillation. Bars show 95% CIs. A rug plot along the *x*-axis shows the distribution of observations.

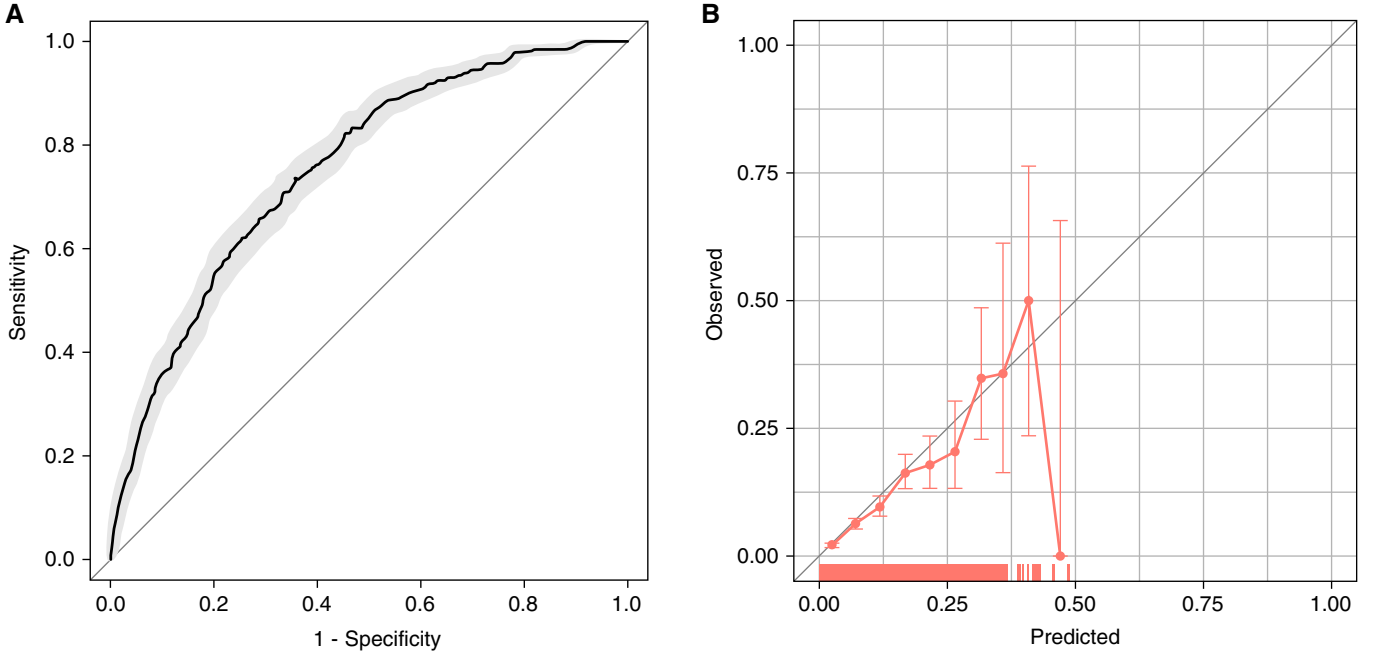


**Figure 2.** Test performance of an updated prediction model for new-onset atrial fibrillation. (*A*) Receiver operating characteristic curve for prediction of new-onset atrial fibrillation in the next 24 hours. Area under the receiver operating characteristic curve = 0.755 (95% confidence interval [CI], 0.734–0.774). (*B*) Calibration curve for prediction of new-onset atrial fibrillation in the next 24 hours. Each circle represents a grouping of patients by predicted probability of atrial fibrillation and the corresponding observed occurrence of atrial fibrillation. Bars show 95% CIs. A rug plot along the *x*-axis shows the distribution of observations.

1.2.1355. This study was designated by the Boston University Institutional Review Board as not human subjects research.

## Results

From 2007 to 2012, 12,304 adult patients were admitted with sepsis, among whom 7,844 had available nurse-charted heart rhythms and no AF before ICU admission. During the first 7 days of admission, 554 (7%) patients developed new AF. Values for pertinent covariates at time of ICU admission, compared with the reference study, are shown in Table 1. Applying the original model to the Medical Information Mart for Intensive Care III cohort demonstrated weak discrimination (C-statistic = 0.598; 95% confidence interval [CI], 0.587–0.609), poor goodness of fit (modified HL chi-square = 6,847; $P < 0.001$), and ICI of 0.15 for new-onset AF (Figure 1). PPV of the model determined at the optimal sensitivity and specificity (Youden's index 0.13) was 7.6% (95% CI, 7.3–8.1%).

Similar results were found in sensitivity analyses using a complete case cohort ($n = 3,633$; C-statistic = 0.580; 95% CI, 0.562–0.598; modified HL chi-square = 956; $P < 0.001$; ICI = 0.03; PPV = 7%; 95% CI, 6.8–7.8%) and limiting to patients aged ⩾40 years ($n = 7,225$; C-statistic = 0.566; 95% CI, 0.553–0.577; modified HL chi-square = 6,842; $P < 0.001$; ICI = 0.16; PPV = 7.9%; 95% CI, 7.5–8.3%). We found similar model performance in medical, surgical, and cardiac ICUs, with the ranges of performance across subgroups demonstrating C-statistic: 0.571–0.638, modified HL $P < 0.001$ (in all subgroups), ICI: 0.12–0.16, PPV: 6.7–10.6%.

We then assessed a revised model using the same covariates but updated intercept and β-estimates in the test cohort. Notable changes in the updated model included the variables of inflammation and immunosuppression no longer showing strong associations with new-onset AF, and reversal in the direction of effect of duration of ICU stay. Evaluation of this model in the validation cohort showed improvements in performance (C-statistic = 0.755; 95% CI, 0.734–0.774; modified HL chi-square = 17; $P = 0.033$; ICI = 0.01; PPV = 10.7%; 95% CI, 9.1–13.9%) (Figure 2).

## Discussion

We validated a risk score designed to predict development of AF during sepsis in an external cohort. Both model discrimination (from C-statistic 0.8 in the original study to 0.598 in the current cohort) and goodness of fit (HL chi-square test 9.6 to 6,847) worsened markedly when applying an unmodified model to an external validation sample. However, model performance improved modestly (C-statistic = 0.755, HL test = 17) after revising the model estimates for performance in the new cohort, though all models had low PPV for new-onset AF. These findings have general ramifications for prediction model development and validation in the ICU setting and AF prediction in particular.

Our findings provide further examples of poor model performance in the ICU setting when applying prediction models to external cohorts (14–16) and show the importance of rigorous external validation before implementation of risk prediction models. In the context of the current study, loss of predictive validity may be due to differences in case mix, small sample size in the initial derivation cohort, or differences in the rate of the outcome of interest. The proportion of patients who developed AF (7%) in our cohort was lower than the 23% reported in the reference study (5) but consistent with prior literature (1, 3) showing an incidence of new AF among 6–10% of ICU patients with sepsis.

Although external validation of the original AF prediction model in a new cohort yielded poor discrimination and goodness of fit, improved performance of a revised model showed that statistical adaptation of models to a new context can be feasible. Although the calibration of the updated model is improved overall (ICI 0.01 vs. 0.15 in our initial model), the worsening calibration at higher risk levels (e.g., >50% risk) seen in the calibration plot suggests that the model may require further improvements to provide reliable enrichment of clinical trials for patients at high risk for new-onset AF. Our findings highlight the patient factors that demonstrate consistent association with new-onset AF across ICU settings, and building models that include these elements while adding other types of data (17) may further enhance risk prediction.

There are important strengths to this study. It was based on a large cohort of 7,844 patients and more than six times the number of observations in the original derivation cohort, with comparable inclusion criteria and similar variable definitions as the original model. Potential limitations include data from a single U.S. academic center, missing data, and claims-based definitions of sepsis. However, sensitivity analyses confirmed the robustness of our findings across different means of handling missingness.

## Conclusions

A previously designed prediction model that predicted daily risk of new-onset AF among ICU patients with sepsis did not perform well in an external validation cohort. Further research is needed to design tools that effectively predict AF occurrence across diverse cohorts of patients with sepsis. ■

**Author disclosures** are available with the text of this letter at www.atsjournals.org.

Justin M. Rucci, M.D.*
Nicholas A. Bosch, M.D.
*Boston University School of Medicine*
*Boston, Massachusetts*

Emily K. Quinn, M.A.
*Boston University School of Public Health*
*Boston, Massachusetts*

Ki H. Chon, Ph.D.
*University of Connecticut*
*Storrs, Connecticut*

David D. McManus, M.D.
*University of Massachusetts Medical School*
*Worcester, Massachusetts*

Allan J. Walkey, M.D.
*Boston University School of Medicine*
*Boston, Massachusetts*

ORCID ID: 0000-0002-6400-6060 (J.M.R.).

*Corresponding author (e-mail: justin.rucci@bmc.org).

## References

1 Walkey AJ, Wiener RS, Ghobrial JM, Curtis LH, Benjamin EJ. Incident stroke and mortality associated with new-onset atrial fibrillation in patients hospitalized with severe sepsis. *JAMA* 2011;306:2248–2254.

2 Meierhenrich R, Steinhilber E, Eggermann C, Weiss M, Voglic S, Bögelein D, *et al.* Incidence and prognostic impact of new-onset atrial fibrillation in patients with septic shock: a prospective observational study. *Crit Care* 2010;14:R108.
3 Walkey AJ, Greiner MA, Heckbert SR, Jensen PN, Piccini JP, Sinner MF, *et al.* Atrial fibrillation among Medicare beneficiaries hospitalized with sepsis: incidence and risk factors. *Am Heart J* 2013;165:949–955.e3.
4 Walkey AJ, Hammill BG, Curtis LH, Benjamin EJ. Long-term outcomes following development of new-onset atrial fibrillation during sepsis. *Chest* 2014;146:1187–1195.
5 Klein Klouwenberg PM, Frencken JF, Kuipers S, Ong DS, Peelen LM, van Vught LA, *et al.*; MARS Consortium*. Incidence, predictors, and outcomes of new-onset atrial fibrillation in critically ill patients with sepsis. A cohort study. *Am J Respir Crit Care Med* 2017;195: 205–211.
6 Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1–W73.
7 Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, *et al.* MIMIC-III, a freely accessible critical care database. *Sci Data* 2016;3:160035.
8 Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit Care Med* 2001;29:1303–1310.
9 Ding EY, Albuquerque D, Winter M, Binici S, Piche J, Bashar SK, *et al.* Novel method of atrial fibrillation case identification and burden estimation using the MIMIC-III electronic health data set. *J Intensive Care Med* 2019;34:851–857.
10 Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. *Ann Transl Med* 2016;4:30.
11 Nattino G, Pennell ML, Lemeshow S. Assessing the goodness of fit of logistic regression models in large samples: a modification of the Hosmer-Lemeshow test. *Biometrics* 2020;76:549–560.
12 Austin PC, Steyerberg EW. The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. *Stat Med* 2019;38:4051–4065.
13 Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32–35.
14 Damluji A, Colantuoni E, Mendez-Tellez PA, Sevransky JE, Fan E, Shanholtz C, *et al.* Short-term mortality prediction for acute lung injury patients: external validation of the Acute Respiratory Distress Syndrome Network prediction model. *Crit Care Med* 2011;39:1023–1028.
15 Görges M, Peters C, Murthy S, Pi S, Kissoon N. External validation of the "quick" Pediatric Logistic Organ Dysfunction-2 score using a large North American cohort of critically ill children with suspected infection. *Pediatr Crit Care Med* 2018;19:1114–1119.
16 Witteveen E, Wieske L, Sommers J, Spijkstra JJ, de Waard MC, Endeman H, *et al.* Early prediction of intensive care unit-acquired weakness: a multicenter external validation study. *J Intensive Care Med* 2020;35:595–605.
17 Bashar SK, Ding EY, Walkey AJ, McManus DD, Chon KH. Atrial fibrillation prediction from critically ill sepsis patients. *Biosensors (Basel)* 2021;11:269.





## Radiologic Classification of Black Lung: Time for a New Gold Standard?

*To the Editor*:

As experienced B-readers, we read with interest the publication by Friedman and colleagues (1). The authors identified a strong association between radiograph classification and history of payment by employer or claimant. The analysis is impressive and interesting, and at first glance, the discordance between readers is quite concerning. However, the following points are notable.

1. According to Table E1, 23,689 (62%) of 37,530 miners had only one reading, and a further 6,057 (16%) had multiple readings with complete agreement on classification. Thus, the system seems to have worked appropriately in almost 80% of black lung applicants. The remaining 7,784 (21%) had multiple readings with disagreement. These miners had a total of 25,315 readings or an average of 3.3 readings per miner (compared with 2.4 readings per miner in those with concordant readings). This difference in the number of readings will magnify the discordance between readers.
2. Because these discordant cases were presumably contested, it is possible that some concordant reads that did not fit the desired narrative would not have been included in the claim and would not have been part of the record when searched.
3. About 50% of the B-readers were nonradiologists and perhaps less experienced with digital imaging and postprocessing, which may have led to over- or under-reading of abnormality.
4. Two separate tasks are inherent in radiographic interpretation for pneumoconiosis: perception of the abnormality and determination of whether the perceived abnormality is consistent with pneumoconiosis. There are no clear guidelines for the latter subjective decision. The 2020 revised version of the classification form (2) may remove some of this ambiguity by asking whether there are any classifiable parenchymal abnormalities. However, this change makes it more important to identify alternate causes from the clinical history.
5. The discussion indicates that "when looking only at B-readers who read almost exclusively in one direction (99% of cases), there were three times more B-readers providing eight times more classifications among those affiliated with employers compared with those affiliated with miners." This likely reflects the greater resources of employers to request and pay for multiple B-readings from physicians, an asymmetry that is likely to increase the number of discordant reads.
6. Most importantly, there is no gold standard for diagnostic truth. In place of the current unhelpful adversarial competition between positive and negative readings, it would probably be less costly and more efficient to acquire a volumetric high-resolution computed tomography (HRCT) for contested cases, interpreted by an approved panel of readers who follow the standards set by the International